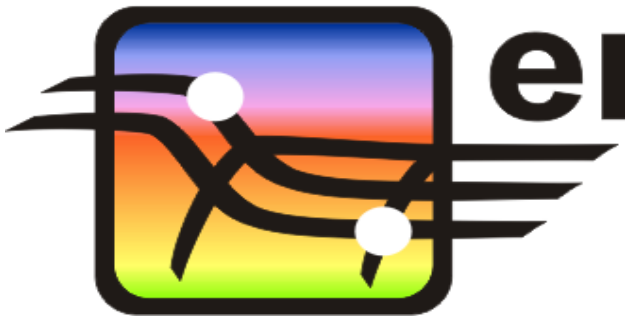


enλIGHTened computing

Applications-Driven Adaptive Compute, Instrument, and Network Resources Integration

Yufeng Xin

MCNC

RTP, NC USA

Sep. 7, 2006

NSF seed funded project

Participating institutes and senior personnel

- **MCNC:** Gigi Karmous-Edwards (PI), John Moore, Steve Thorpe, Lina Battestilli, Bonnie Hurst, Mark Johnson, Yufeng Xin.
- **Louisiana State University:** Ed Seidel (PI), Gabrielle Allen, Seung-Jong Park (Jay), Andrei Hutanu, Tevfik Kosar, Jon MacLaren.
- **Renaissance Computing Institute (RENCI):** Prof. Dan Reed (PI), Lavanya Ramakrishnan.
- **North Carolina State University:** Prof. Harry Perros (PI).
- **Partners**:
  - Cisco, Calient, AT&T Research, and IBM
  - Other research projects and initiatives: NLR, Dragon, Cheetah, SURA
  - International partners: Glambda, PHOSPHORUS, and GLIF.

Outline

- **Enlightened overview**
  - Motivation and methodology
  - Testbed
  - Software System Architecture
- **Extended network service provisioning**
  - Temporal and spatial extension
  - Control and management plane integration
  - Integrated resource allocation and fault tolerance
  - Middleware interface
- Preliminary implementation and experiment

Motivations

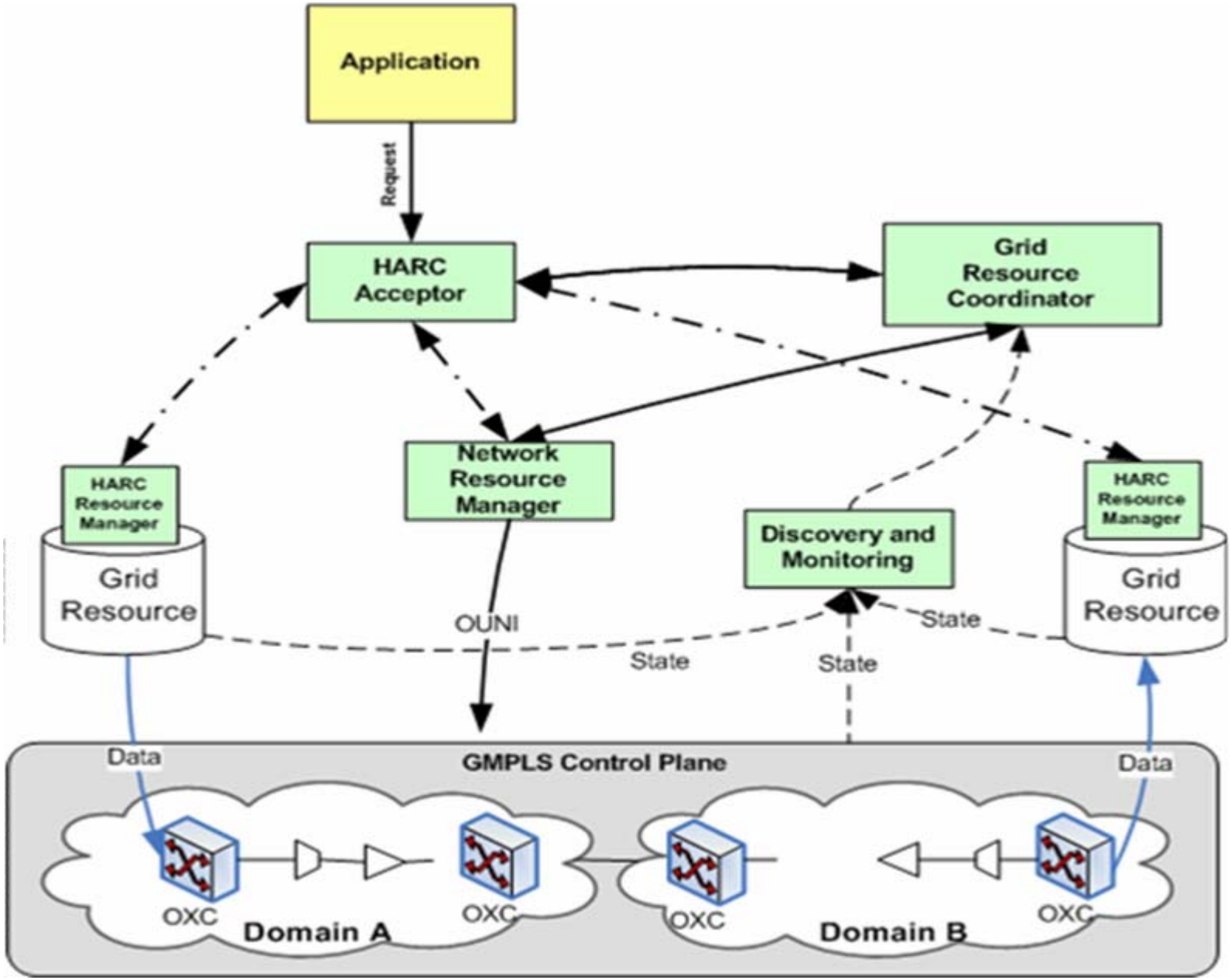
- Ubiquitous and efficient utilization of the distributed scientific facilities
- The need for dynamic high-capacity end-to-end circuits
- The need for the integrated services to optimally allocate and control compute, storage, instrument, and networking resources
- Control, management, and middleware plane integration
  - Scalability
  - Hierarchical network service provisioning

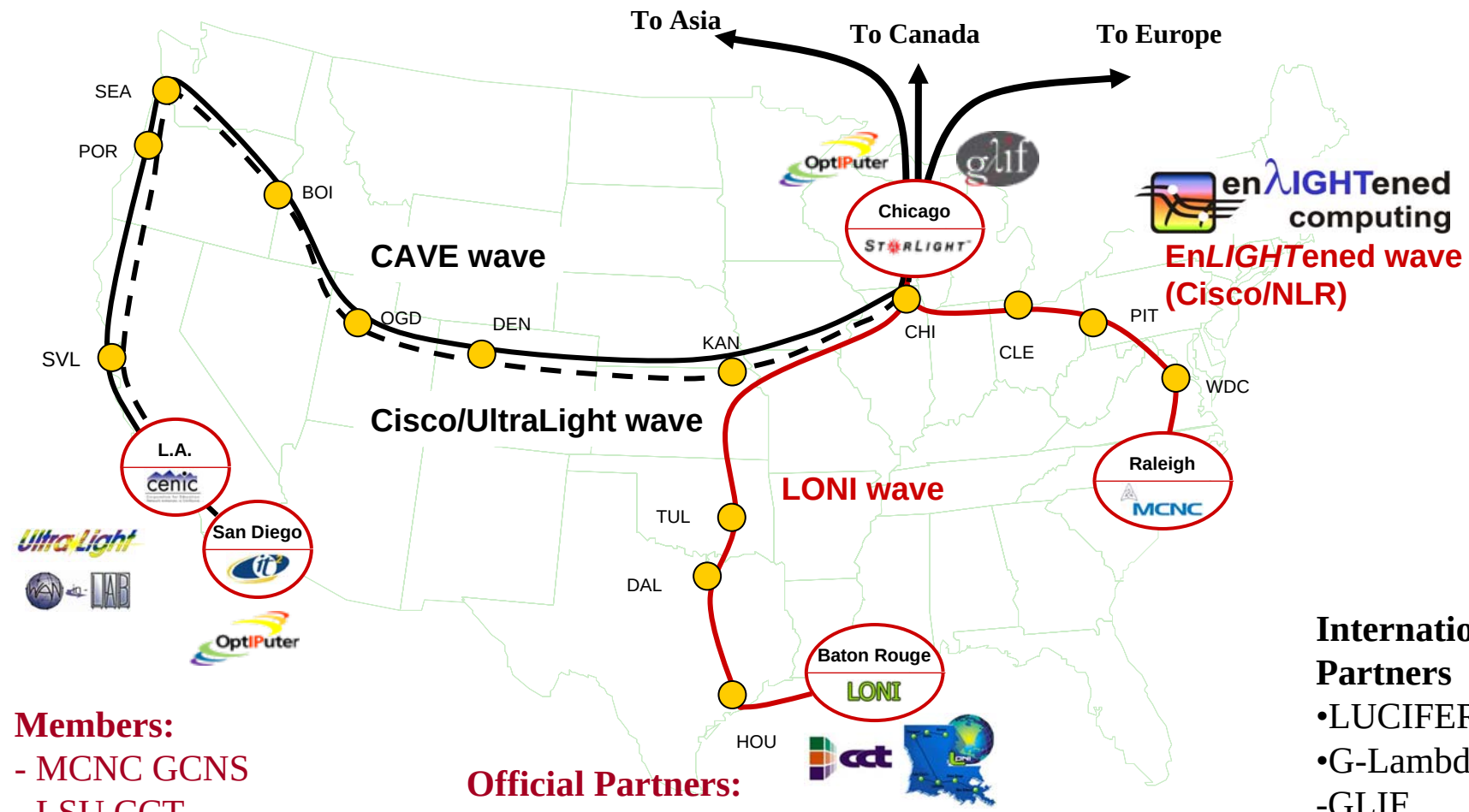
R&D challenges

- The need to standardize the interfaces among Grid middleware and the network.
- Coordination and Co-scheduling of Network resources with other Grid resources (CPU, databases, sensors, instruments)
- Discovery and monitoring -based system-level feedback control
- Extended L1/2 network services
  - On-demand vs. In-advance
  - Unicast, multicast, and anycast
- Control, management and middleware plane integration
  - GMPLS networking
  - Reconfiguration and re-optimization
  - Application controlled networking via the Grid middleware
- Testbed enabling dynamic service provisioning
  - GMPLS enabled PXC, Ethernet switch….
  - E-NNI

System level methodology and architecture

- Testbed peering: meaningful scale
  - Starlight, Ultralight, Loni wave, JGN-II…
  - GLIF
- System peering
  - GLambda, Japan
  - PHOSPHORUS, EU
- Vertical integration via monitoring-based feedback control
  - Application abstraction layer
  - Resource management layer
  - Service layer
  - Resource layer


Application
Request
HARC Acceptor
Grid Resource Coordinator
Network Resource Manager
HARC Resource Manager
Grid Resource
Discovery and Monitoring
HARC Resource Manager
Grid Resource
OUNI
State
State
State
Data
GMPLS Control Plane
Data
OXC
Domain A
OXC
OXC
Domain B
OXC

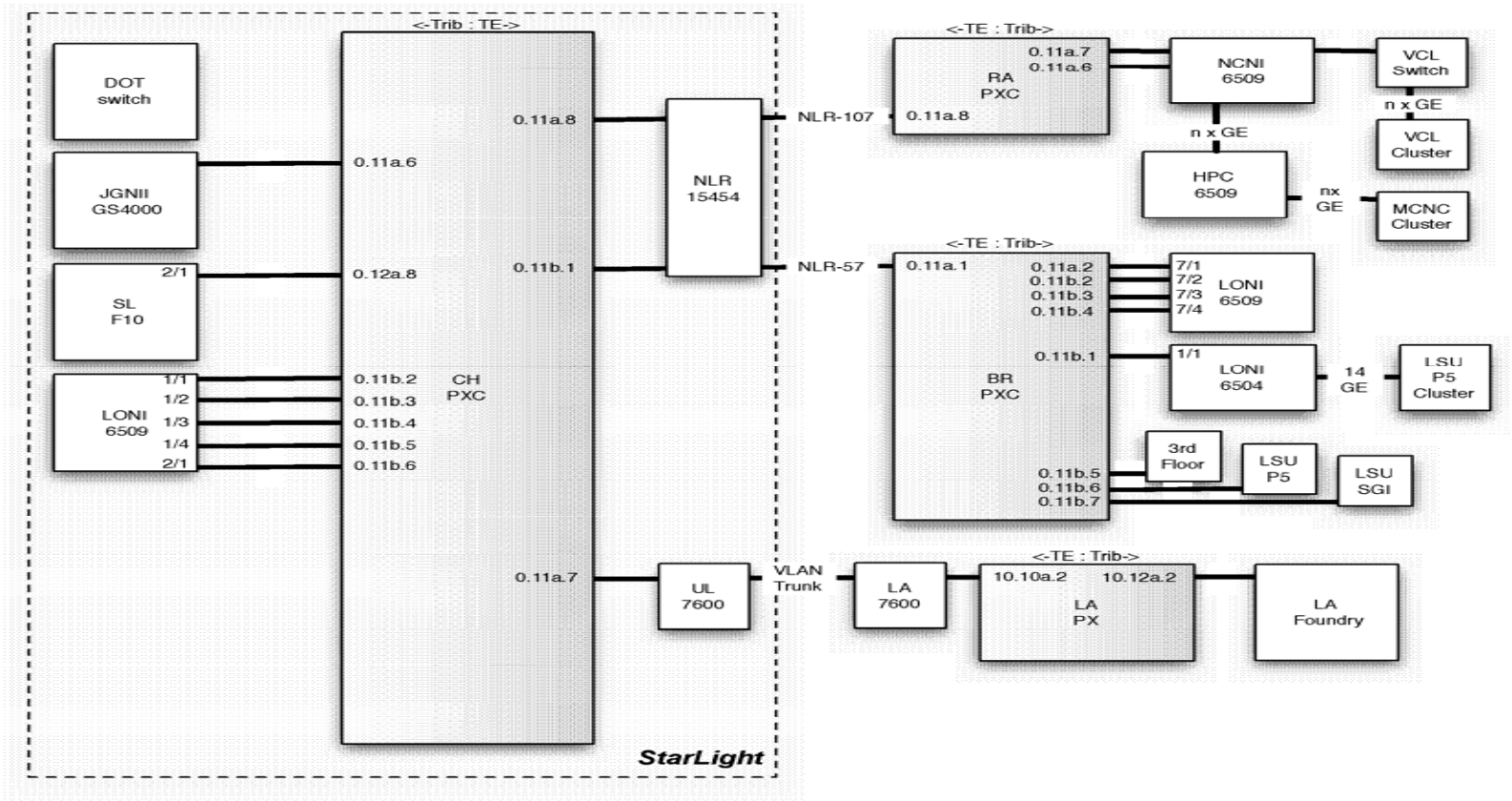


Members:
- MCNC GCNS
- LSU CCT
-NCSU
-(Subcontract) RENCI

Official Partners:
- AT&T Research
- SURA
- NRL
- Cisco Systems
- Calient Networks
- IBM

NSF Project Partners
- OptIPuter
- UltraLight
- WAN-in-LAB
- DRAGON
- Cheetah

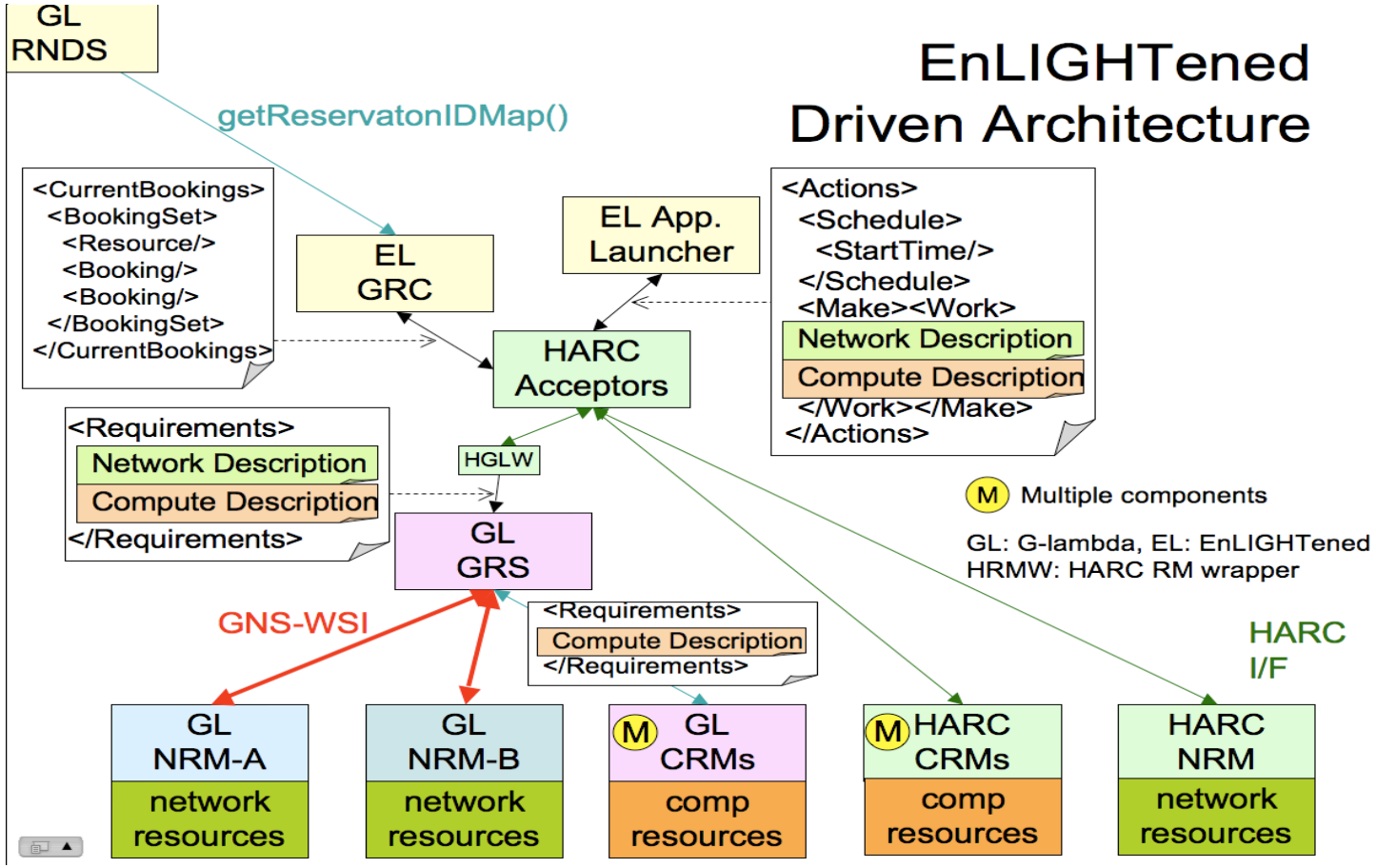
International Partners
•LUCIFER - EC
•G-Lambda - Japan
-GLIF



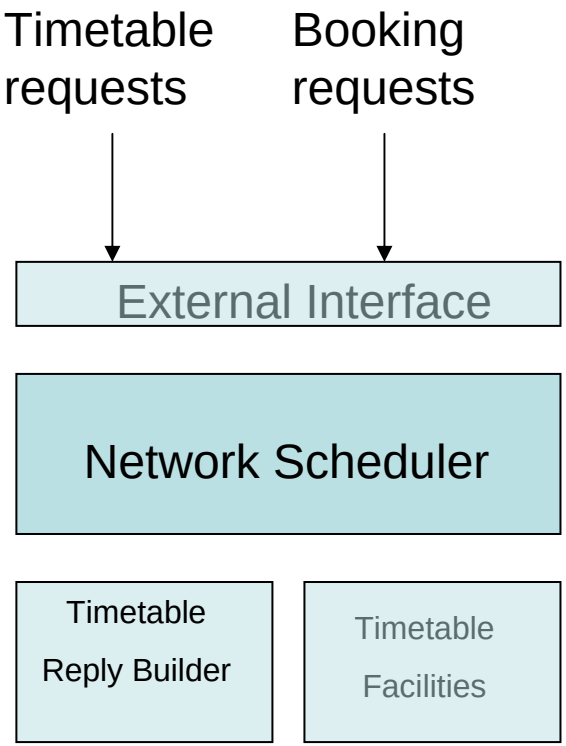
Enlightened Testbed
v0.6
6/13/06
jhm@mcnc.org

Web Service based implementation and experiment

- Highly-Available Robust Co-Scheduler (HARC) solves the distributed transaction problem. (LSU: Jon Mcleren)



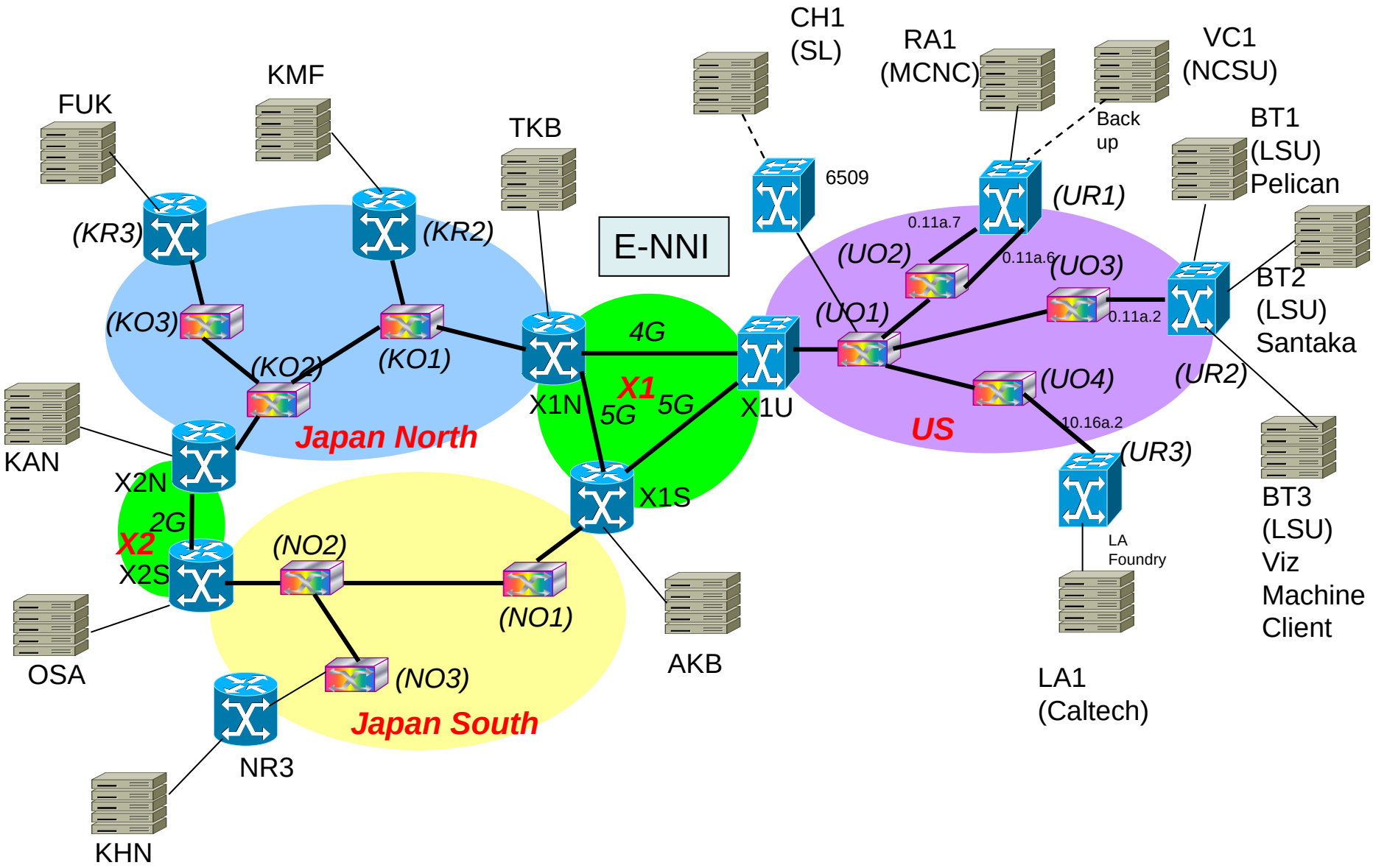
NRM design



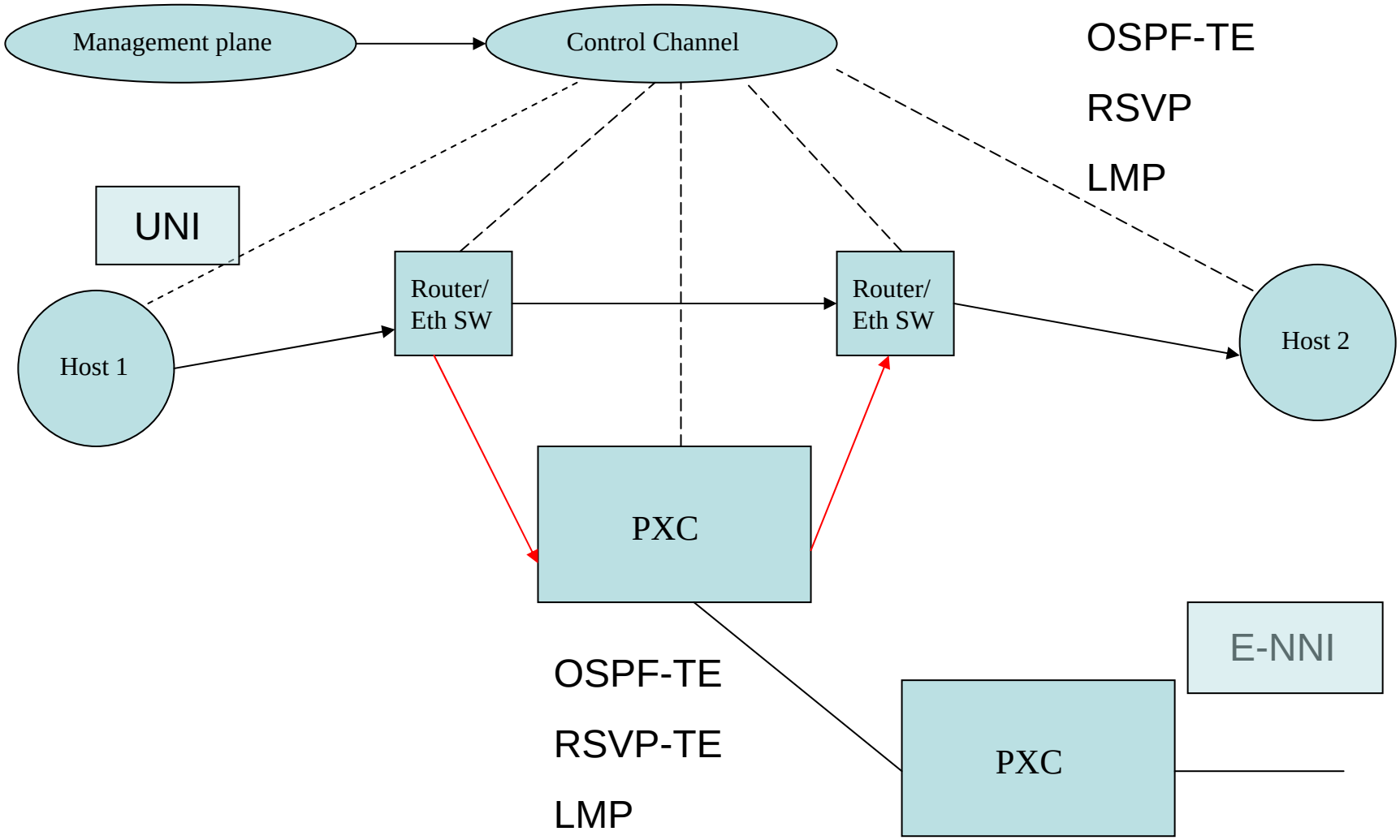
- Look up the ***Path***, find a list of links of the path (source, destination).
- Check the max bandwidth for each link. The request is ``Aborted'' if the required bandwidth is not supported by the links.
- If bandwidth is ok, then look at the ***Timetable*** for each link in the list.
- If all links are available for the requested interval:
  - Mark the links as reserved in the timetables
  - Return the status as ``Prepared''.

- Pre-computed path for node pairs
- Update according to the monitoring information
- Reservation via GMPLS

Internetworking experiment



Work-in-Progress: GMPLS/MPLS across domains



Network management and provisioning service

- Extended network provisioning service (E-NPS)
- Control and management plane integration
- Network performance monitoring
- Integrated service provisioning and fault tolerance

Network provisioning service (NPS): Algorithm study and design

- A single path: **SinglePath**(source, destination(s), bandwidth, QoS_Attributes, Time_Attributes)
  - Unicast
  - Anycast
  - Shared

- A number of paths allocated at the same time frame: **GroupPath**(<SinglePath>)

  Virtual topology

- Multicast connection: **Multicast**(source, <destination1,...,destination2>, QoS_Attributes, Time_Attributes)

Management and control plane design

- Questions
  - In-advance reservation (GMPLS fails)
  - On-demand reservation starvation (unfair in-advance allocation)
  - Multi-granularity connection management (100M->1 GE->10GE, GMPLS stack)

- Multi-time-scale network resource management/control
  - L3/2/1 reconfiguration
  - Service provisioning re-optimization
  - Co-provisioning with other resource
  - Resource discovering and performance monitoring: Mona-Lisa

Integrated fault management

- Fault handling mechanisms
  - Fail-stop: stop the application;
  - Ignore the failure: continue the application execution;
  - **Fail-over**: assign the application to new resources and restart;
  - **Migration**: replication and reliable group communication to continue the execution.
- Fault recovery
  - Fail-over or migrate within the same host(s):
  - Fail-over or mitigate to different host(s)

Conclusion

- Multi-layer architecture and team formation
- National footprint testbed with GMPLS support
- National and international partnership
- Balanced research, development, and experiment efforts

www.EnlightenedComputing.org

Thank You !!!

Welcome to GLIF06 Demo

Sep. 11~13

Inter-domain advance reservation of coordinated network and computing resources over the Pacific